Comparing potential recharge estimates from three Land Surface Models across the Western US

REWATI NIRAULA

Department of Hydrology and Atmospheric Sciences, University of Arizona, Tucson, Arizona

THOMAS MEIXNER

Department of Hydrology and Atmospheric Sciences, University of Arizona, Tucson, Arizona

HOORI AJAMI

Department of Environmental Sciences, University of California Riverside, Riverside

MATTHEW RODELL

Hydrological Sciences Branch, NASA Goddard Space Flight Center, Greenbelt, Maryland

DAVID GOCHIS

NCAR HR Regional Modelling, Boulder, Colorado

CHRISTOPHER L. CASTRO

Department of Hydrology and Atmospheric Sciences, University of Arizona, Tucson, Arizona

Corresponding author address: Rewati Niraula, Department of Hydrology and Atmospheric Sciences, University of Arizona, 1133 E James E Rogers Way, JW Harshbarger Bldg. Rm 202, Tucson, AZ 85719, USA. E-mail: rewatin@email.arizona.edu

Abstract:

Groundwater is a major source of water in the western US. However, there are limited recharge estimates available in this region due to the complexity of recharge processes and the challenge of direct observations. Land surface Models (LSMs) could be a valuable tool for estimating current recharge and projecting changes due to future climate change. In this study, simulations of three LSMs (Noah, Mosaic and VIC) obtained from the North American Land Data Assimilation System (NLDAS-2) are used to estimate potential recharge in the western US. Modeled recharge was compared with published recharge estimates for several aquifers in the region. Annual recharge to precipitation ratios across the study basins varied from 0.01-15% for Mosaic, 3.2-42% for Noah, and 6.7-31.8% for VIC simulations. Mosaic consistently underestimates recharge across all basins. Noah captures recharge reasonably well in wetter basins, but overestimates it in drier basins. VIC slightly overestimates recharge in drier basins and slightly underestimates it for wetter basins. While the average annual recharge values vary among the models, the models were consistent in identifying high and low recharge areas in the region. Models agree in seasonality of recharge occurring dominantly during the spring across the region. Overall, our results highlight that LSMs have the potential to capture the spatial and temporal patterns as well as seasonality of recharge at large scales. Therefore, LSMs (specifically VIC and Noah) can be used as a tool for estimating future recharge rates in data limited regions.

1. Introduction

Groundwater is a life-sustaining natural resource that supplies water to billions of people on earth (Gleeson et al. 2012). It plays a central part in irrigated agriculture and sustaining

43 ecosystems (Giordano 2009; Siebert et al 2010), and enables human adaptation to climate
44 variability and change (Taylor et al. 2012). Globally it accounts for 1/3rd of all fresh-water
45 withdrawals, for domestic (36%), agricultural (42%) and industrial purposes (27%) (Doll et al.
46 2012; Taylor et al. 2012). In the United States (US), ground water is the source of drinking
47 water for 50% of the population and as much as 90% of the population in rural areas, especially
48 in the West (Anderson and Woosley 2005). Reduced reliability of surface water supplies in the
49 western US with projected increases in evaporative demand and uncertain changes in annual
50 precipitation (Rasmussen et al. 2011, 2014) may increase groundwater use (Scanlon 2005).
51 Many areas of the region are already experiencing groundwater depletion caused by sustained
52 groundwater pumping (Faunt 2009; Konikow 2013; Castle et al. 2014).
53 Groundwater recharge is a flux of water into the saturated zone. Spatial variability of
54 recharge rates is controlled by precipitation and other climate variables (Hoekstra and Meknnen
55 2012; Hoekstra et al. 2012; Gleeson et al. 2012), vegetation, soils, and geology (Stonestrom et al.
56 2007). Despite the importance of groundwater in this region, limited recharge estimates are
57 available due to the complexity of recharge processes and the lack of feasible measurement
58 methods (Scanlon et al. 2006). Thus, improving current recharge estimates and understanding
59 spatial variability of recharge processes are essential for sustainably managing this precious
60 resource (Scanlon et al. 2006; Famiglietti and Rodell 2013) to meet human and ecosystem
61 demands in the future (Scanlon et al. 2006).
62 Recharge estimation methods include water balance accounting, remote sensing,
63 observational methods and environmental tracer analysis, and modeling (Scanlon et al 2006;
64 Healy, 2010). In the western US, groundwater recharge generally occurs at depth, where direct
65 observational methods cannot be applied. Several Land Surface Models (LSMs) (e.g SAC-SMA

66 (Burnash et al. 1973; Burnash 1995; SSiB, Xue et al. 1991); Mosaic (Koster and Suarez 1996);
67 NSSIP (Koster et al. 2000); Variable Infiltration Capacity (VIC; Liang et al. 1994) ; Noah (Ek et
68 al. 2003, Nui et al. 2011); and CLM (Bonan et al. 2002, 2011)) have been developed over the
69 last few decades to better represent land surface and atmospheric processes as well as improve
70 estimates of various water, energy and carbon fluxes at the land surface. These models could be a
71 valuable tool for estimating current and future recharge estimates due to projected climate
72 change. However, to date, besides currently published recharge estimates (Li et al. 2015) in the
73 eastern US using the VIC model, recharge estimates from these models have not been
74 comprehensively assessed.

75 LSMs vary in representation of the exchange of energy, mass, momentum and CO_2
76 exchange between the land surface and the overlying atmosphere (Koster and Suarez 1996;
77 Liang et al. 1994; Bonan et al. 2011; Nui et al. 2011). It is therefore important to understand how
78 differences in model structure affect the simulation of recharge and whether certain LSMs
79 perform better under particular physiographic and climatic settings. In this paper, we compare
80 recharge estimates from three LSMs over the western US with a specific emphasis on 10 aquifer
81 systems where recharge estimates from other approaches are available.

82 The major questions addressed in this study are: 1. Are recharge estimates in the western
83 US from various LSMs significantly different? 2. Do LSMs provide reasonable estimates of
84 recharge in the western US? 3. Do the amount, seasonality, trend and spatial pattern of recharge
85 vary based on the choice of LSMs? For addressing these questions, simulations of three LSMs
86 (Mosaic, Noah, and VIC) obtained from the North American Land Data Assimilation System-
87 phase 2 (NLDAS-2) were used for assessing recharge estimates across the western US. We used
88 MODIS ET (Mu et al. 2011) and baseflow index (BFI) based recharge (Wolock 2003a) for the

whole western US for comparison and evaluation purpose. Simulated recharge from the LSMs was compared with published recharge estimates from 10 aquifers in the region (Northern Plains, Central High Plains, Southern High Plains, San Pedro, Death Valley, Salt Lake Valley, Central Valley, Columbia Plateau, Spokane Valley, and Williston, Fig 1) synthesized by Meixner et al. (2016). These aquifers represent a broad sample of variability in climatological, geological, and hydrological characteristics along with anthropogenic pressures like groundwater pumping on the aquifers. Trends, amounts, and patterns of recharge from the three models were compared statistically to determine their consistency. Statistical analyses (Kolmogorov-Smirnov test, Kendall Tau trend analysis test, spatial pattern correlation test) were conducted using R (version R 3.1.3).

2. Methods:

2.1. North American Land Data Assimilation System Phase 2 (NLDAS-2):

NLDAS-2 (Mitchell et al. 2004; Xia et al. 2012) integrates observation-based and model reanalysis data to drive LSMs offline. It executes 4? LSMs at 1/8th-degree grid spacing at an hourly temporal scale over central North America, enabled by the Land Information System (LIS; Kumar et al. 2006; Peters-Lidard et al. 2007). LIS is a scalable land data assimilation system that integrates a suite of advanced LSMs, high resolution satellite and observational data, data assimilation and parameter optimization techniques, and high-performance computing tools (Kumar et al. 2006; Peters-Lidard et al. 2007). Outputs from three LSMs (Mosaic, Noah, and VIC) over a 30-year historical period (1981-2010) were used to answer the study questions posed above. The first two years of the simulation (1979-1980) were used as a model spin up period and excluded from the analysis. The data used in this study were generated within NASA's Earth

Science Division and are archived and distributed by the Goddard Earth Sciences (GES) Data and Information Services Center (DISC; http://disc.gsfc.nasa.gov/hydrology/index.shtml).

Past NLDAS-2 multi-model evaluations are performed for evaluating surface water and energy fluxes (Wei et al 2013), such as soil moisture (Mo et al. 2011; Mo et al. 2012; Xia et al. 2014; Xia et al. 2015a; Xia et al. 2015b), evapotranspiration (Long et al. 2014), soil temperature (Xia et al. 2013), and streamflow (Mo et al. 2012; Xia et al. 2012b). Although there have been efforts to compare the continental scale water and energy-fluxes among LSMs for NLDAS-2, (Xia et al. 2012a) model performance was only evaluated with surface water (Xia et al. 2012b). No particular efforts have been made using these models to simulate and characterize the amount and seasonality of recharge in the western US, mostly due to the limited availability of recharge data. However, recently Li et al. (2015) demonstrated that NLDAS/VIC provides reasonable estimates of groundwater recharge in the central and northeastern US.

2.2 Descriptions of LSMs

2.2.1 Formulation of Water and Energy Budget in LSMs:

Although the 3 LSMs vary in complexity for the treatment of exchange of energy, mass, momentum and CO_2 between land surface and overlying atmosphere, they follow similar fundamental conceptualization of the energy and water budget. The water balance is calculated based on the continuity equation:

$$ds/dt = P-ET-R-G$$

where, ds/dt is the change in strorage (mm), P is precipitation (mm), R is surface runoff (mm), G is ground water runoff (mm) and ET is evapotranspiration (mm) which is calculated as,

$$ET= CE+BE+T+S$$

where, *CE* is canopy evaporation (mm), *BE* is bare soil evaporation (mm), *T* is transpiration (mm) and *S* is sublimation (mm).

The models assume gravity-driven, free-drainage from the bottom layer as subsurface runoff/recharge, and surface runoff is the excess water after infiltration.

The surface energy balance is calculated based on the equation:

$$R_n = LE + SH + G + SF$$

where, R_n is the net radiation flux (W/m^2), *LE* (λET) is the latent heat flux (W/m^2), *SH* is the sensible heat flux (W/m^2), *G* is the ground heat flux (W/m^2), and *SF* is the snow phase-change heat flux (W/m^2).

2.2.2. Mosaic

The 1D Mosaic LSM (Koster and Suarez 1992; Koster and Suarez 1996, Table 1) which calculates both energy and water balance, accounts for subgrid heterogeneity of land surface characteristics by dividing each grid cell into several subregions, called ‘‘tiles,’’. Each tile contains a single vegetation or bare soil type (Koster and Suaraz, 1996). Energy and water balance calculations are performed separately over each tile and weighted by fractional coverage to calculate the total fluxes for each grid cell. The vertical structure of the model includes a single canopy layer and three soil layers: a thin surface layer (0-10cm), a middle layer (10-40cm) that encompasses the remainder of the root zone, and a lower ‘‘recharge’’ layer (40-200cm) at the bottom. Mosaic calculates total evapotranspiration as the sum of bare soil evaporation, transpiration, and canopy evaporation. Runoff occur both as overland flow during precipitation events and as delayed baseflow. Mosaic treats baseflow/recharge as a linear function of water content, bedrock slope, and hydraulic conductivity of the bottom layer.

2.2.3. Noah

157 Noah (Ek et al. 2003, Table 1) is a 1-D column model that simulates soil moisture, soil
158 temperature, snow depth, snow water equivalent, canopy water content, and water and energy
159 flux terms of the surface water and energy balance (Mitchell 2004). For this study, Noah was
160 configured to have a 2-m-deep soil layer divided into the 4 sub-layers:0-10 cm, 10-40 cm, 40-
161 100 cm, and 100 cm -200 cm. The deepest layer acts as a reservoir with gravity drainage at the
162 bottom. The volumetric soil moisture is determined using the diffusive form of Richard's
163 equation. The total evaporation, in the absence of snow, is the sum of direct evaporation from the
164 topmost soil layer, evaporation of precipitation intercepted by plant canopy, and transpiration
165 from the vegetation canopy. The Noah LSM assumes spatially continuous soil moisture values
166 within tiles pixels, parameterizes surface runoff with a simple infiltration-excess scheme, and
167 treats base- flow/recharge as a linear function of hydraulic conductivity (K) of the bottom soil
168 layer (Schaake et al. 1996).

169 **2.2.4. VIC**

170 The VIC (Liang et al. 1994, Table 1) model incorporated within NLDAS-2, and characterizes the
171 subsurface as three soil layers with spatially variable thickness. The surface is described by
172 different vegetation types plus bare soil. The land cover types are specified by their leaf area
173 index (LAI), canopy resistance and relative fraction of roots in each of the soil layers.
174 Evapotranspiration from each vegetation type is characterized by potential evapotranspiration
175 together with canopy resistance and aerodynamic resistance to water transfers. Associated with
176 each land cover type are a single canopy layer, and multiple soil layers (up to 2 m depth).
177 Subsurface hydrology parameterizations of the VIC LSM is more complex (Liang et al. 1994)
178 because it uses a spatial probability distribution to represent subgrid heterogeneity in soil
179 moisture. It also treats baseflow/recharge as a nonlinear recession curve which is a function of

bottom layer soil moisture (it is linear below a threshold and then non-linear above that threshold). The top 2 soil layers are designed to represent the dynamic behavior of the soil column that responds to rainfall events and evapotranspiration, and the lower layers control inter-storm soil moisture behavior. The lower layer only responds to rainfall when the upper layer is fully saturated and thus can separate subsurface flow from quick response storm flow. Roots can extend down to the bottom layer, depending on the vegetation and soil type. The soil characteristics (such as soil texture, hydraulic conductivity, etc.) are held constant for each grid cell. In the model, all the states and output variables are calculated for each land cover tile at each time step and weighted by fractional land cover to calculate the total fluxes for each grid cell.

2.3 **Descriptions of study area/basins:**

2.3.1. Western US:

The Western US (Fig 1) is the largest region of the country, covering more than half of the land area of the contiguous US. It is also the most geographically diverse region in the country encompassing the Pacific Coast, the temperate rainforests of the Northwest, the highest mountain ranges (including the Rocky Mountains, the Sierra Nevada, and Cascade Range), the Great Plains, and all of the desert areas located in the US (the Mojave, Sonoran, Great Basin, and Chihuahua deserts). Elevation varies between -86 m to 4402 m above sea level (Fig 1).

The Western US consists primarily of five land-use/land-cover classes: grassland/shrubland (59%), forest (28.1%), agriculture (6.3%), developed (1.5%), and barren (1.9%) (Sleeter at al. 2012). Grassland/shrubland and barren lands are most common in the arid-southwest and interior desert regions, whereas forest dominates in the Pacific Northwest and

202 Rocky Mountains. Agriculture and developed areas are found to some degree in nearly all
203 regions but are concentrated mainly in a relatively few high-density areas (USGS 2012).
204 As a generalization, the climate of the Western US can be described as overall semiarid;
205 however, parts of the region get extremely high amounts of rain and/or snow, and other parts are
206 true desert and get little rain per year. Annual rainfall (Fig 2) ranges between 58 mm to 5051 mm
207 based on NLDAS 2 data and is greater in the eastern portions, gradually decreasing until
208 reaching the Pacific Coast where it again increases.

209 **2.3.2. Study Basins**

210 **High Plains Aquifer**

211 The High Plains aquifer (HPA, Fig 1) extends into eight States: Colorado, Kansas, Nebraska,
212 New Mexico, Oklahoma, South Dakota, Texas and Wyoming. The aquifer is comprised of
213 unconsolidated, poorly sorted clay, silt, sand and gravel and is underlain by bedrock units. HPA
214 is divided into Northern (NHP), Central (CHP) and Southern (SHP) regions. Average
215 precipitation (P) in HPA is 522 mm/yr and average recharge is 48 mm/yr (Meixner et al 2015).
216 In general, average annual P and recharge increase from south to north (Table 2) and occurs
217 predominantly during summer.

218 **San Pedro Aquifer**

219 The San Pedro Basin (Fig 1) in southern Arizona is representative of the hydrogeology of a
220 southern Basin and Range aquifer system (Goode and Maddock 2000). It is an alluvial aquifer
221 that is comprised of basin-bounding crystalline and sedimentary rock mountains and
222 unconsolidated sediments of clay, silt, sand, and gravel within the valley (Pool and Dickinson
223 2006). The basin receives an annual average precipitation and recharge of 400 mm and 6.5 mm
224 respectively (Table 2). The majority of annual rainfall (~54%) in the San Pedro occurs during the

summer monsoon season, with the remainder occurring in the winter months as rain and snow from low-intensity storms.

Death Valley Aquifer

The Death Valley Aquifer System (Fig 1) is located in the arid southern Great Basin of Nevada and California. Major aquifers consist of fractured volcanic rock and alluvium. The average annual P and recharge for the aquifer is 185 mm and 2.8 mm respectively (Table 2). Precipitation particularly as snowfallin mountain systems is predominantly in the winter months.

Salt Lake Valley (SLV) Aquifer

SLV aquifer (Fig 1), a representative of northern Basin and Range aquifer system consists of shallow unconfined aquifers underlain by confined to semi-confined sand and gravel aquifers, (Lambert 1995; Cederberg et al. 2009). The average annual P and recharge for the aquifer are 488 mm and 203 mm respectively (Table 2) with most of the P falling as snow during winter and spring.

Central Valley Aquifer

The Central Valley aquifer system of California (Fig 1) is an unconsolidated sand and gravel aquifer that underlies the Sacramento and San Joaquin Valleys of central California. The average annual P and recharge for the aquifer are 650 mm and 315 mm respectively (Table 2). About 85% of the precipitation falls from November to April in Central Valley.

Columbia Plateau Aquifer

The Columbia Plateau aquifer system (Fig 1) in Washington, Oregon, and Idaho (Kahle et al. 2011) consists of productive basalt aquifers characterized by highly permeable interflow zones separated by less permeable basalt-flow interiors. Extensive sedimentary aquifers consisting of

valley-fill deposits lie atop the basalts along major drainages. With an average annual P of 440 mm mostly occurring during winter months, recharge is estimated to be 162 mm (Table 2).

Spokane Valley-Rathdrum Prairie Glacial Aquifer

The Spokane Valley-Rathdrum Prairie aquifer (Fig 1) is a glacial aquifer in northwestern Idaho and northeastern Washington (Hutson et al. 2004). The aquifer is composed of coarse-grained sediments with fine-grained layers interspersed (Hsieh et al. 2007; Kahle and Bartolino 2007). Annual average P is 689 mm/yr and mostly is concentrated during winter, and average recharge is 300 mm/yr (Table 2).

Williston Basin Glacial Aquifer System

The Williston Basin (Fig 1) is present within southern Canada, northeastern Montana, and western North Dakota (Soller et al. 2012). The aquifer is composed of till, clay, silt, sand, and gravels (Fullerton et al. 2004). The average P and recharge estimates are 382 mm/yr and 4.7 mm/yr respectively (Table 2). Both P and recharge are summer dominated.

2.4. Evaluation Datasets:

2.4.1 MODIS Evapotranspiration data:

The MOD16 (Mu et al 2011) global evapotranspiration (ET) datasets (Fig 3) are regular 1-km^2 land surface ET datasets for the global vegetated land areas at 8-day, monthly and annual intervals. The dataset covers the time period 2000 – 2010. The ET algorithm is based on the Penman-Monteith equation (Monteith 1965). Terrestrial ET includes evaporation from wet and moist soil, from rain water intercepted by the canopy before it reaches the ground, and the transpiration through stomata on plant leaves and stems. Evaporation of water intercepted by the canopy is a very important water flux for ecosystems with a high LAI. It should be noted that the MODIS ET estimates used for evaluation have lots of uncertainties associated with input data

(e.g. LAI, PAR), inaccuracy of measured ET data from eddy covariance flux towers, scaling from flux towers to landscape, and algorithms (associated with processes and parameters). However, it is the best available dataset for the region due to its spatial and temporal coverage.

2.4.2 Baseflow Index (BFI) recharge:

A spatially distributed recharge map (Wolock 2003a, Fig 5) was created by multiplying a grid of base-flow index (BFI) values (Wolock 2003b) by a grid of streamflow values (Gebert et al. 1987) derived from a 1951-1980 mean annual runoff contour map generated for the whole USA. The assumptions inherent in this recharge estimation approach are that in long term: (1) recharge is equal to discharge, and (2) the BFI reasonably represents the proportion of natural ground-water discharge to streamflow. The BFI grid (1 km resolution) was interpolated from the BFI values of 8,249 U.S. Geological Survey stream gages (Wolock 2003c) using the inverse distance weighting interpolation method. The BFI values are computed using an automated hydrograph separation computer program called the BFI program (Wahl and Wahl 1988; Wahl and Wahl 1995). However it should be noted that the BFI-based recharge itself is a very rough estimate and should not be treated as an observation due to high uncertainty related to this dataset. The recharge dataset likely reflects general patterns across broad geographic regions, but recharge values at specific sites are unlikely to be accurate.

2.4.3 Basin wide literature estimates of recharge:

The literature based recharge estimates for the study basins synthesized by Meixner et al. (2016) are used for evaluating model estimates. These recharge estimates come from various sources and different approaches (observational, environmental tracer analysis, and modeling) are used

for making these estimates. Although the methodology is not consistent over the study basins, these estimates are the best available.

3. Results and Discussions:

3.1. Comparing ET: Among models and with MODIS ET

The models tend to agree on the spatial pattern of ET (Table 3, Fig 3) with each other and with the MODIS ET following the pattern of P (Fig 2), though ET rates vary. MODIS ET was generally lower than LSMs ET. Mosaic consistently generated higher ET compared to Noah and VIC. Model ensemble mean slightly improved the spatial pattern of ET when compared with MODIS ET (Fig 3).

ET more or less followed the pattern of P across the region (Fig 2, Fig 3). Annual average precipitation ranges between 58 mm to 5051 mm based on NLDAS 2 data (Fig 2). A gradual decrease in P from east towards west before a significant increase in P at the west coast was observed. Among the basins examined, Death Valley ($\overline{P}$ = 185 mm) and Spokane Valley ($\overline{P}$ = 689 mm) are the driest and the wettest basins respectively (Table 1). Average annual ET was estimated between 58 mm and 1260 mm for Mosaic, between 36 mm and 1123 mm for Noah, between 21 mm and 986 mm for VIC and between 49 mm and 1026 mm based on model ensemble mean. The models and MODIS ET (Fig 3, Table 3) showed the lowest ET in southern regions and higher ET in western coast and lower eastern regions (Fig 3). Higher ET on the western coast is related to higher water availability from higher P. Higher ET in south eastern corner was due to the combined effect of T and P (high T and moderate P).

MODIS estimated ET (ranged between 35 mm and 1175 mm) was generally lower than LSMs ET over the western US (Fig 3). Annual ET is the highest for the Mosaic LSM model, and lowest for MODIS. LSM's ET was higher than the MODIS ET for most of the study basins. Mean annual ET values across the study basins were between 176 mm and 597 mm (87% and 99% of P) based on Mosaic, 120 mm and 454 mm (59% and 91% of P) based on Noah, 153 mm and 485 mm (54% and 89% of P) based on VIC and 170 mm and 490 mm (71 and 92% of P) based on MODIS (Table 4). The Spokane valley has the lowest ET/P ratio according to all models (Table 5). The San Pedro basin has the highest ET/P ratio based on Mosaic and Noah, and Central High Plains (CHP) based on VIC estimates.

Mosaic consistently generated higher ET compared to Noah and VIC for most of the Western US (Fig 4). Overall, using the MODIS estimates as the standard, Mosaic overestimated ET by 36% (Fig 4). Noah and VIC follow a similar pattern with overestimation for low ET areas and underestimation for higher ET areas (Fig 3), but overall the bias was minimum (4% for Noah and 9% for VIC). Model ensemble mean overestimated MODIS ET by 15%. It should be noted that MODIS based ET is based on an empirical retrieval algorithm that requires calibration using ?data from a network of in situ measurements which themselves require calibration. MODIS area averaged ET rates have not yet been proven to be more accurate than LSMs, hence the biases shown here should not be interpreted as errors.

The specific breakdown of ET differs among models with Mosaic more dominated by passive processes and the other two more vegetative processes. Over the western US, Mosaic generated most of the ET through bare soil evaporation (47%) followed by transpiration (33%), canopy evaporation (18%) and sublimation (2%). Noah produced most of the ET through transpiration (41%) and bare soil evaporation (39%), followed by canopy evaporation (16%) and

sublimation (4%). VIC on the other hand generated a majority of its ET as transpiration (82%) with other contributions from canopy (13%) and sublimation (4%). The very high contribution of ET through transpiration and limited contribution from bare soil by VIC is related to the tiling process in VIC which classifies a majority of land areas to some vegetation group. Moreover, the root zone depth extends throughout the 2m soil layer in VIC, while root zone depth is up to 1 m in the case of Mosaic and Noah (excluding forest land cover in Noah).

Relatively larger magnitudes of ET by Mosaic compared to other models could be ascribed to greater upward diffusion of water from deeper soil layers to the shallow root zone (Mitchell et al. 2004; Long et al. 2013). This process has a significant influence on the recharge estimates because the recharge in the LSMs is the function of water content in the bottom soil layer. These differences could be related to energy balance or water balance constraints (Mitchell et al. 2004; Peters-Lidard et al. 2011; Rodell et al. 2011; Xia et al. 2012b; Cai et al. 2014).

3.2. Comparing Recharge across the Western US: Among models and with BFI-based recharge

A similar spatial pattern (Table 6, Fig 4) of recharge was observed based on LSMs and also with BFI, although recharge rates vary among models as in the case of ET. While Mosaic consistently generated lower recharge compared to Noah, VIC and BFI, VIC overestimated recharge at low recharge zones and underestimated recharge in medium to higher recharge zones compared to Noah. Noah and VIC models overestimated recharge compared to BFI.

Average annual recharge rates varied between 0 and 4128 mm based on Noah, 0 and 3479 mm based on Mosaic, 0 and 2209 mm based on VIC, 0 and 3272 mm based on model ensemble and between 0 and 2031 mm based on BFI- based estimates at 1/8 degrees grid scale (Fig 4). The average recharge rates for the whole western US was estimated to be 139 mm based

359 on Noah, 46 based on Mosaic, 123 based on VIC, 103 mm based on mean model ensemble and
360 82 mm based on BFI.

361 Although recharge rates differ among LSMs and BFI-based estimates, high and low
362 recharge zones are similar among them (Fig 4). It was observed that BFI-based recharge
363 captures higher recharge zones for the west coast, and it predicts lower rates for the Eastern US
364 compared to LSMs (Fig 4). Mosaic showed slightly different patterns from BFI- recharge in
365 other regions except the west coast mostly because Mosaic generated lower recharge compared
366 to Noah, VIC and BFI (Fig 5). While Mosaic consistently generated lower recharge compared to
367 Noah, VIC overestimated recharge at low recharge zones and underestimated recharge in
368 medium to higher recharge zones compared to Noah (Fig 5). There was a stronger relationship
369 between Noah and BFI-based recharge (R^2:0.76) and model ensemble (R2= compared to VIC
370 (R^2:0.62) and Mosaic (R^2:0.56) with BFI-based recharge (Fig 6?). Results from a Kolmogorov-
371 Smirnov (K-S) test suggested that the recharge estimates from the three models as well as the
372 ensemble average are significantly different from each other and from the BFI-based recharge
373 estimates (Table 7). Spatial pattern and magnitude of Model mean were more similar to VIC and
374 Noah model and did not necessarily better than those models when compared with the BFI based
375 recharge (Fig 4). Model estimates differ less in the higher recharge zones in the east and west
376 coast and more in the inner dry regions (Fig 4).

377 It should be noted that the BFI-based recharge itself is a very rough estimate and should
378 not be treated as an observation due to high uncertainty related to this dataset. The dataset is
379 likely to underestimate natural recharge in arid regions where ET is significant. Also, ground-
380 water discharge to streams does not occur in "losing" streams which are more common in arid

regions (Wolock 2003c). As a result, the BFI-based recharge consistently underestimated recharge compared to literature estimates in the study basins.

3.3. Evaluating LSMs's recharge estimates at a basin scale:

Mosaic consistently underestimated recharge across all the basins where estimates are available. Noah captured recharge reasonably well in wetter basins in xx, but overestimated it in drier basins (xx and xx). VIC overestimated recharge in the drier basins and underestimated it for wetter aquifers. Over the study basins, recharge estimates varied between 0.2 mm/yr to 97.6 mm/yr based on the Mosaic model, between 12.4 mm/yr to 289.6 mm/yr based on the Noah model, and between 22.5 mm/yr to 201.7 mm/yr based on the VIC model (Table 4 & Fig 6). The literature based recharge estimates ranged between 2.8 mm/yr and 315.5 mm/yr (Table1, Fig 6). Model ensemble mean ranged between 14.1 mm/yr and 189.1 mm/yr (Fig 6) and was not ? necessarily better than Noah and VIC when evaluated with the literature estimates Ensemble mean is affected by highly skewed lower Mosaic estimates (Fig 6). Although models have predicted different recharge rates for the study basins, the patterns of predicted recharge were similar. Models agree in identifying drier and wetter aquifers (i.e. low vs. high recharge aquifers, Fig 6) although the driest and wettest aquifers identified by the models varied slightly among each other. The driest and wettest aquifers were the Death Valley aquifer and the Central Valley aquifer respectively based on the literature estimates (Fig 6). Similar results were obtained for the VIC model. However, Mosaic and Noah produced different results.

Based on literature estimates for individual aquifers in the region, about 1% to 49% of the precipitation becomes recharge (Table 5), lowest for the Williston basin and highest for the Central Valley. The Williston basin is in a semi-arid region, and thus the fraction of precipitation, which occurs primarily during summer that becomes recharge is relatively small.

Recharge in the Central Valley comes from irrigation return flows, diffuse recharge directly from precipitation and from mountain system recharge in the form of leakage from streams originating in the Sierra Nevada Mountains. These basins where a higher proportion of P becomes recharge are more permeable and have lower ET rates.

Based on Mosaic, about 0.01% to 15% of precipitation becomes recharge in the study basins which is much lower than literature estimates (Table 5). Based on Noah, about 3.2% to 42% of P becomes recharge in the study basins which is within similar range compared to literature estimates (Table 5). Based on VIC, about 6.7% to 31.8% of the precipitation becomes recharge in the study basins which is slightly higher for the drier basin and slightly lower for the wetter basins when compared to literature estimates. Models agreed with the literatures estimates that basins like Spokane valley, Central Valley and Columbia produce higher percent of P as recharge (Table 5).

Overall, Mosaic consistently underestimated recharge significantly across the basins (Fig 6). VIC slightly overestimated recharge in the drier basins (Death Valley, Williston basin, San Pedro basin, SHP and CHP) but slightly underestimated in wetter basins (NHP, Colombia, SLK, Spokane Valley and Central Valley, Fig 6). Noah, on the other hand overestimated recharge in the drier basins but capture recharge reasonably well in the wetter basins except for SLV where it underestimated recharge (Fig 6). Thus, based on the analysis of these 10 basins, although none of the models were found to be capturing the recharge magnitude across the whole western US, it can be said that the Noah model showed a great promise in capturing the recharge in wetter regions. Mosaic seems to work better in drier basins which could just be an artifact that it underestimates recharge throughout the region. VIC seems to balance between Noah and Mosaic and seems to work for both dry and wetter regions if a single model is to be chosen across the

western US. However overall, all three models (especially VIC and Noah) showed a lot of promise that with some advancements/ modifications in hydrologic process representation and with some calibration at local scale/aquifer these models can be a useful tool for estimating current recharge and also for forecasting the effect of projected climate change on recharge.

3.4. Seasonality of Recharge:

The models were fairly consistent with respect to the seasonality of recharge, which was largest during the spring. Over the study basins, models tend to agree on the seasonality of recharge occurring dominantly during spring months (MAM) except in the SHP basin (Fig 7). This spring time dominance of recharge is mostly pronounced in Mosaic. Since VIC has a more damped response to recharge, it produced similar recharge throughout the year for many basins. Seasonality of recharge did not necessarily follow the seasonality of precipitation in the aquifers (Fig 7).

Seasonally higher recharge in spring for the basins could be due to additional sources of snowmelt from winter (DJF) P which tends to melt at the beginning of spring when temperature is sufficient to melt but not high enough to lose a lot of water from evaporation (Dunne and Leopold 1978; Clark and Fritz 1997, Ajami et al. 2012; Jasechko et al. 2014) in addition to rain occurring in spring time. Several field monitoring studies in Sweden (Rodhe 1981), Idaho (Flerchinger et al. 1992), and the United States mid-west (Delin et al. 2007; Dripps 2012) have also found that the spring snowmelt constitutes the bulk of annual groundwater recharge at the middle latitudes examined here.

3.2.2. Potential reasons for differences in recharge estimates among models:

Differences in recharge estimates among models can be attributed to differences in (1) ET calculations/estimates, (2) model structure particularly the thickness of bottom soil layer and (3) parameterizations. Recharge estimates from Mosaic were significantly smaller than those of Noah and VIC (Figs. 4, 5, 6). The lower estimates of recharge by Mosaic were directly related to Mosaic's very high estimated ET. The model converted most of the precipitation to evaporation leaving much less water available to run off or infiltrate and percolate down as recharge. All of these LSMs characterize ET using soil moisture stress factor that impacts evaporation from the top soil layer and vegetation transpiration. As noted previously, the Noah model in NLDAS-2 has four soil layers with spatially invariant thicknesses of 10, 30, 60, and 100 cm. The first three layers form the root zone in non-forested regions, with the fourth layer included in forest regions. The Mosaic model has three layers with thicknesses of 10, 30, and 160 cm with the first two layers corresponding to the root zone. Mosaic has a greater ability to transfer water from the deep layer to the surface/root zone through vertical diffusion, and therefore shows higher ET rates under normal conditions (Long et al. 2013). This process dries up the bottom soil layer and leaves minimal water to become recharge through drainage. Although vertical diffusion does occur in Noah, the magnitude is much smaller compared to Mosaic. No vertical diffusion between 3rd and 2nd layer occurs in VIC model which also accounts for the sub-grid heterogeneity of vegetation and soil moisture. In VIC, rooting depth extends to the bottom layer unlike Noah and Mosaic. Although the general conceptualization and basic structure of the models are similar, they vary in certain processes and formulations. These differences in the parameterizations can give rise to large variability in the outputs depending upon the variables of interest. The multi-model analysis carried out under the Global Soil Wetness Project-2 (GSWP-2) (Dirmeyer et al. 2006) illustrated

473 that LSM variables, especially those associated with snow processes (i.e., snow water equivalent)
474 and soil water (i.e., soil moisture in the lower layers), have a large spread. The same is true for
475 groundwater recharge (Xia et al. 2012a).

476 The thickness of the bottom layer, which is the source of recharge, is 160 cm in Mosaic
477 model, 100 cm in Noah model, and of variable depth in VIC model. This variability leads to
478 different amounts of free drainage. It is likely that differences in soil wetness, related to
479 evapotranspiration and surface runoff rates, have a greater impact on the modeled recharge
480 estimates than the free drainage formulations themselves (ref?). The relative bias analysis of soil
481 moisture in the US showed that the models have small relative biases for the Eastern US where
482 soils are normally wet but large relative biases in the western region where soils are drier (Xia et
483 al. 2014). The disparity in mean annual evaporation and runoff ratio among the LSMs was also
484 most obvious over the western mountainous regions (Xia et al. 2012a).

485 **3.6. Limitations of LSMs:**

486 Like most LSMs, those used in this study, were developed using many simplifications
487 necessary to represent complex physical processes across large spatial scales with limited
488 computational power and with imperfect inputs. These LSMs have soil columns with depths of 2
489 m, divided into multiple (3 or 4) layers, while neglecting deeper soil moisture and groundwater.
490 Vertical flows of soil water are estimated using the Richards equation while the horizontal flow
491 is ignored. Groundwater recharge is parameterized by a gravitational percolation term, which is a
492 linear/nonlinear function of bottom soil layer drainage affected by soil type, soil moisture
493 content, and slope. It derives from a simple infiltration/saturation excess scheme used for both
494 surface runoff and drainage. None of the models take into account the horizontal flow of
495 groundwater. The partitioning of saturation excess into surface runoff and drainage and how they

vary in space are also quite different from one LSM to another (Lohmann et al. 1998, 2004; Boone et al. 2004). Nevertheless, LSMs provide spatially and temporally continuous estimates of hydrological variables that would be impossible to obtain using observations alone, and often the results are surprisingly good considering their limitations (Dirmeyer et al. 2006; Syed et al. 2008; Jimenez et al. 2011; Wang et al. 2011; Li et al. 2015).

4. Summary and Conclusions:

Three LSMs: Mosaic, Noah and VIC were used to estimate recharge and assess its spatial pattern and temporal trend in the western US. While Mosaic estimates were consistently low compared to the BFI based recharge, Noah recharge estimates were generally higher. VIC has mixed results with higher estimates at lower recharge zones and lower estimates at high recharge zones when compared with the BFI based recharge. Models were consistent in identifying high and low recharge zones although rates vary. When evaluated with published estimates of recharge in 10 aquifers across the western US, Mosaic was consistent in underestimating recharge significantly across all the basins. VIC slightly overestimated recharge in the dry aquifers and slightly underestimated it in the wetter aquifers. Noah captured recharge reasonably well for wetter basins (SHP, NHP, Colombia, Spokane and Central Valley), but overestimated it in the other basins. The models accurately identified low and high recharge aquifers, although their rankings based on recharge magnitude differed. The models were fairly consistent with respect to the recharge seasonality, which was largest during the spring. VIC's recharge seasonality was dampened compared to Noah and Mosaic. This consistency among models was greater in the south than in the north, with its more snow dominated regions.

518 Overall, LSMs have the potential to capture the spatial and temporal patterns, as well as
519 seasonality of recharge across the western US. Mosaic in particular requires calibration to
520 capture the magnitude of recharge. Noah is more useful in capturing recharge in wetter regions
521 with default parameters and VIC could be useful for both drier and wetter conditions but might
522 require some calibration to improve estimations. In general, all three models (especially VIC
523 and Noah) showed promise that with advancements/ modifications in hydrologic process
524 representation and with additional calibration at local/aquifer scale, these models can be a very
525 useful tool for estimating current recharge and also for forecasting the effect of projected climate
526 change on recharge.
527 Even though the source of meteorological forcing data produced as part of NLDAS-2 for
528 all these LSMs was the same, differences in recharge estimates among models emerged due to
529 differences in ET calculations/estimates, model structure particularly the thickness of the bottom
530 layer, and parameterizations. Calibration of these LSMs could improve their ability to estimate
531 recharge. However, it should be noted that carefully calibrating LSMs at a regional scale and at a
532 grid level can be computationally and labor-intensive and observational data for calibration are
533 also limited. Improving model inputs and adding process complexity especially associated with
534 groundwater mechanisms in future could help reduce uncertainty in recharge estimates. Recharge
535 estimates were highly controlled by precipitation and there was not much of an imprint of
536 topography on the recharge estimates, even for major mountain chains, given that such patterns
537 are evident in ET and likely in precipitation.

538 **Acknowledgments:** We would like to acknowledge USGS John Wesley Powell Center for
539 funding this research. We also express our sincere gratitude to all the members of Powell Center
540 working group (Andrew H. Manning, David A. Stonestrom, Diana M. Allen, Kyle W. Blasch,

Andrea E. Brookfield, Jordan F. Clark, Alan L. Flint, Kirstin L. Neff, Bridget R. Scanlon, Kamini Singha, and Michelle A. Walvoord) for their valuable suggestions/research directions.

References:

Ajami, H., T. Meixner, F. Dominguez, J. Hogan, and T. Maddock, 2012: Seasonalizing Mountain System Recharge in Semi-Arid Basins-Climate Change Impacts. *Groundwater*, **50**, 585-597. doi: 10.1111/j.1745-6584.2011.00881.x

Anderson, M. T. and L. H. Woosley, 2005: Water availability for the Western United States—Key scientific challenges: U.S. Geological Survey Circular 1261, 85 p.

Bonan, G. B., 1996: A land surface model (LSM version 1.0) for ecological, hydrological, and atmospheric studies: Technical description and user's guide, NCAR Tech. Note NCAR/TN-417+STR, 150 pp., Natl. Cent. for Atmos. Res., Boulder, Colo.

Bonan, G. B., K. W. Oleson, M. Vertenstein, and S. Levis, 2012: The land surface climatology of the community land model coupled to the NCAR community climate model. Journal of Climate, 15, 3123-3149.

Bonan G. B, P. J. Lawrence, K. W. Oleson, S. Levis, M. Jung, M. Reichstein, D. M. Lawrence, and S. C. Swenson, 2011: Improving canopy processes in the Community Land Model version 4 (CLM4) using global flux fields empirically inferred from FLUXNET data. *Journal of Geophysical Research-Biogeosciences,* **116**, G02014, doi:10.1029/2010JG001593

Boone A. and Coauthors, 2004: The Rhone-Aggregation Land Surface Scheme intercomparison project: An overview. *J. Climate,* **17**, 187–208.

Burnash, R. J. C., 1995: The NWS river forecast system: Catchment modeling. Computer Models of Watershed Hydrology, V. P. Singh, Ed., Water Resources Publications, Littleton, Colorado, 311–366. 188

Burnash, R. J. C., R. L. Ferral and R. A. McGuire, 1973: A generalized streamflow simulation system: Conceptual modeling for digital computers. Tech. rep., U.S. Dept. of Commerce, National Weather Service, Silver Springs, M.D., and State of California, Dept. of Water Resources, Sacramento, Calif.

Cai, X., Z. L. Yang, Y. Xia, M. Huang, H. Wei, L. R. Leung, and M. B. Ek, 2014: Assessment of simulated water balance from Noah, Noah-MP, CLM, and VIC over CONUS using the NLDAS test bed. *J. Geophys. Res. Atmos.,* **119**, 13,751–13,770, doi:10.1002/2014JD022113.

582 Castle, S. L., B. F. Thomas, J. T. Reager, M. Rodell, S. C. Swenson, and J. S. Famiglietti, 2014:
583 Groundwater depletion during drought threatens future water security of the Colorado River
584 Basin, *Geophys. Res. Lett.,* **41**, doi:10.1002/2014GL061055.
585
586 Cederberg, J. R., P. M. Gardner, and S. A.Thiros, 2009: Hydrology of Northern Utah Valley,
587 Utah County, Utah, 1975–2005: U.S. Geological Survey Scientific Investigations Report 2008–
588 5197, 114 p.
589
590 Clark, I. D., and P. Fritz, 1997: Environmental Isotopes in Hydrogeology, A. F. Lewis, N. Y.
591 Delin, G. N., R. W. Healy, D. L. Lorenz, and J. R. Nimmo (2007), Comparison of local-to
592 regional-scale estimates of ground-water recharge in Minnesota, USA, *J. Hydrol.,* **334**, 231–249.
593
594 Dirmeyer, P. A., X. Gao, M. Zhao, Z. Guo, T. Oki, and N. Hanasaki, 2006: GSWP-2:
595 Multimodel analysis and implications for our perception of the land surface. *Bull. Amer. Meteor.*
596 *Soc.,* **87**, 1381–1397
597
598 Döll, P., H. Hoffmann-Dobrev, F.T. Portmanna, S. Siebert , A. Eicker, M. Rodell , G. Strassberg,
599 B.R. Scanlon, 2012: Impact of water withdrawals from groundwater and surface water on
600 continental water storage variations. *J. Geodyn*. **59–60**, 143–156.
601
602 Dripps, W. R., 2012: An integrated field assessment of groundwater recharge. *Open Hydrol.*
603 *J.,* **6**, 15–22.
604
605 Dunne, T., and L. B. Leopold, 1978: Water in environmental planning, New York, 818 pp.
606
607 Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D.
608 Tarpley, 2003: Implementation of Noah land surface model advancements in the National
609 Centers for Environmental Prediction operational mesoscale Eta model. *J. Geophys. Res.,*
610 **108(D22),** 8851, doi:10.1029/2002JD003296.
611
612 Famiglietti, J. S., M. Lo, S. L. Ho, J. Bethune, K. J. Anderson, T. H. Syed, S. C. Swenson, C. R.
613 de Linage, and M. Rodell, 2011: Satellites measure recent rates of groundwater depletion in
614 California's Central Valley. *Geophys. Res. Lett.,* **38**, L03403, doi:10.1029/2010GL046442.
615
616 Famiglietti, J. and M. Rodell, 2013: Water in the balance. Science
617 perspectives, *Science*, **340**, 1300–1301, doi:10.1126/science.1236460.
618
619 Faunt, C. C., Eds., 2009: Groundwater availability of the Central Valley Aquifer, California,
620 U.S. Geol. Surv. Prof. Pap., 1766, 225 pp.
621
622 Faunt, C. C., 2009: Groundwater Availability of the Central Valley Aquifer, California (US
623 Geological Survey, 2009).
624
625 Flerchinger, G. N., K. R. Cooley, and D. R. Ralston (1992), Groundwater response to snowmelt
626 in a mountainous watershed. *J. Hydrol.,* **133**, 293–311.
627

Fullerton, D. S., R. B. Colton, C. A. Bush, and A. W Straub, 2004: Map showing spatial and temporal relations of mountain and continental glaciations on the northern plains, primarily in northern Montana and northwestern North Dakota: U.S. Geological Survey Scientific Investigations Map 2843, http://pubs.usgs.gov/sim/2004/2843/.

Gebert, W. A., D. J. Graczyk, and W.R. Krug, 1987: Average annual runoff in the United States, 1951-80: U.S. Geological Survey Hydrologic Investigations Atlas HA-710, 1 sheet, scale 1:7,500,000.

Giordano, M., 2009: Global groundwater? Issues and solutions. *Annu. Rev. Environ. Resour*. **34**,153–178

Gleeson T. T., Wada Y. Y., M. F. P. Bierkens , and L. P. H., van Beek, 2012: Water balance of global aquifers revealed by groundwater footprint. *Nature* , **488**, 197 200. doi:10.1038/nature11295

Goode, T.C. and T. Maddock III, (2000). Simulation of groundwater conditions in the Upper San Pedro Basin for the evaluation of alternative futures, Arizona Research Laboratory for Riparian Studies, Department of Hydrology and Water Resources, Tucson, Arizona: University of Arizona, 122p.

Healy, R.W., 2010: Estimating Groundwater Recharge, Cambridge: Cambridge University Press. 264 p.

Hoekstra, A. Y. and Mekonnen, M. M., 2012: The water footprint of humanity. *Proc. Natl Acad. Sci*., **109**, 9, 3232-3237, http://dx.doi.org/10.1073/pnas.1109936109.

Hoekstra, A. Y., Mekonnen, M. M., Chapagain, A. K., Mathews, R. E. and B. D. Richter, 2012: Global monthly water scarcity: blue water footprints versus blue water availability. *PLoS ONE*, **7**, e32688

Houston, N.A., S. L. Gonzales-Bradford, A. T. Flynn, S. L. Qi, S. M. Peterson, J. S. Stanton, , D. W. Ryter, T. L. Sohl, and G. B. Senay, 2013: Geodatabase compilation of hydrogeologic, remote sensing, and water-budget-component data for the High Plains aquifer, 2011: U.S. Geological Survey Data Series 777, 12 p.

Hsieh, P.A., Barber, M.E., Contor, B.A., Hossain, Md. A., Johnson, G.S., Jones, J.L., and Wylie, A.H., 2007: Ground-water flow model for the Spokane Valley-Rathdrum Prairie Aquifer, Spokane County, Washington, and Bonner and Kootenai Counties, Idaho: U.S. Geological Survey Scientific Investigations Report 2007-5044, 78 p.

Jasechko, S., S. J. Birks, T. Gleeson, Y. Wada, P. J. Fawcett, Z. D. Sharp, J. J. McDonnell, and J. M. Welker, 2014: The pronounced seasonality of global groundwater recharge, *Water Resour. Res.,* **50**, 8845–8867, doi:10.1002/2014WR015809.

673 Jimenez, C. and Coauthors, 2011: Global inter-comparison of 12 land surface heat flux
674 estimates. *J. Geophys. Res.,* **116**, D02102, doi:10.1029/2010JD014545.
675
676 Kahle, S. C., and J. R. Bartolino, 2007: Hydrogeologic framework and ground-water budget of
677 the Spokane Valley-Rathdrum Prairie aquifer, Spokane County, Washington, and Bonner and
678 Kootenai Counties, Idaho: U.S. Geological Survey Scientific Investigations Report 2007-5041,
679 48 p., 2 pls.
680
681 Kahle, S. C., D. S. Morgan, W. B. Welch, D. M. Ely, S. R. Hinkle, J. J. Vaccaro, and L. L.
682 Orzol, 2011: Hydrogeologic framework and hydrologic budget components of the Columbia
683 Plateau Regional Aquifer System, Washington, Oregon, and Idaho: U.S. Geological Survey
684 Scientific Investigations Report 2011-5124, 66 p.
685
686
687 Konikow, L. F., 2013: Groundwater depletion in the United States (1900-2008): U.S. Geological
688 Survey Scientific Investigations Report 2013-5079, 63 p.
689
690 Koster, R. D., and M. J. Suarez, 1992: Modeling the land surface boundary in climate models as
691 a composite of independent vegetation stands. *J. Geophys. Res.,* **97**, 2697–2715.
692
693 Koster, R. D., and M. J. Suarez, 1996: Energy and water balance calculations in the Mosaic
694 LSM, NASA Tech. Memo., 104606, 9, 1996.
695
696 Koster, R. D., M. J. Suarez, A. Ducharne, M. Stieglitz, and P. Kumar, 2000: A catchment-based
697 approach to modeling land surface processes in a general circulation model: 1. Model structure.
698 *J. Geophys. Res.,* **105**, 24,809–24,822, doi:10.1029/2000JD900327.
699
700 Kumar, S. V., and Coauthors, 2006: Land information system: An interoperable framework for
701 high resolution land surface modeling. Environ. Modell. Software, 21 (10), 1402–1415,
702 doi:10.1016/j.envsoft.2005.07.004.
703
704 Lambert, P. M., 1995: Numerical simulation of ground-water flow in basin-fill material in Salt
705 Lake Valley, Utah: Utah Department of Natural Resources Technical Publication No. 110–B, 58
706 p.
707
708 Li, B., M. Rodell, and J. S. Famiglietti, 2015: Groundwater variability across temporal and
709 spatial scales in the central and northeastern U.S. *J. Hydrology*, accepted.
710
711 Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges, 1994: A Simple hydrologically
712 Based Model of Land Surface Water and Energy Fluxes for GSMs. *J. Geophys. Res.,* **99**(D7),
713 14,415-14,428.
714
715 Liang, X., Z. Xie andM. Huang, 2003: A new parameterization for surface and groundwater
716 interactions and its impact on water budgets with the variable infiltration capacity (VIC) land
717 surface model. *J. Geophys. Res.*, **108**, 8613, doi:10.1029/2002JD003090.
718

719 Lohmann D., and Coauthors, 2004: Streamflow and water balance intercomparison of four land
720 surface models in the North-American Land Data Assimilation System (NLDAS). *Journal of*
721 *Geophysical Research*, **109**, D07S91. DOI: 10.1029/2003JD003571.
722
723 Lohmann, D., and Coauthors, 1998: The Project for Intercomparison of Land-surface
724 Parameterization Schemes (PILPS) phase 2(c) Red-Arkansas River basin experiment: 3. Spatial
725 and temporal analysis of water fluxes. *Global Planet. Change,* **19**, 161–179.
726
727 Long, D., L. Longuevergne, and B. R. Scanlon, 2014: Uncertainty in evapotranspiration from
728 land surface modeling, remote sensing, and GRACE satellites. *Water Resour. Res.,* **50**,1131–
729 1151, doi:10.1002/2013WR014581.
730
731 Meixner T., and Coauthors, 2015: Implications of Projected Climate Change for Groundwater
732 Recharge in the Western United States. *Journal of Hydrology*, **534**, 124-138.
733
734 Mitchell, K. E., and Coauthors, 2004: The multi-institution North American Land Data
735 Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental
736 distributed hydrological modeling system. *J. Geophys. Res.,* **109**, D07S90,
737 doi:10.1029/2003JD003823
738
739 Mo, K. C., L. N. Long, Y. Xia, S. K. Yang, J. E. Schemm, and M. Ek, 2011: Drought indices
740 based on the Climate Forecast System Reanalysis and ensemble NLDAS. *J.*
741 *Hydrometeorol.,* **12(2),** 181–205.
742
743 Mo K. C., L. C. Chen, S. Shukla, T. J. Bohn, D. P. Lettenmaier, 2012: Uncertainties in North
744 American land data assimilation systems over the contiguous United States. *J. Hydromet.,* **13,**
745 996–1009.
746
747 Monteith, J. L., 1965: Evaporation and environment. *Symp. Soc. Exp. Biol.,* **19**, 205–224.
748
749 Mu, Q., M. Zhao, and S. W. Running, 2011: Improvements to a MODIS global terrestrial
750 evapotranspiration algorithm. *Remote Sens. Environ.,* **115(8)**, 1781–1800.
751 doi:10.1016/j.rse.2011.02.019.
752
753 Niu G.-Y., and Coauthors, 2011: The community Noah land surface model with
754 multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale
755 measurements. *J. Geophys. Res*., **116**, D12109, doi:10.1029/2010JD015139.
756
757 Peters-Lidard , C.D. and Coauthors, 2007: High-performance Earth system modeling with
758 NASA/GSFC's Land Information System. *Innovations in Systems and Software*
759 *Engineering*, **3(3)**, 157-165. DOI:10.1007/s11334-007-0028-x
760
761 Pool, D.R. and J. E. Dickinson, 2007: Ground-water flow model of the Sierra Vista
762 Subwatershed and Sonoran portions of the Upper San Pedro Basin, southeastern Arizona, United
763 States, and northern Sonora, Mexico, USGS Scientific Investigations Report 2006-5228, Reston,
764 Virginia: USGS.

765
766 Rasmussen, R., and Coauthors, 2014: Climate Change Impacts on the Water Balance of the
767 Colorado Headwaters: High Resolution Regional Climate Model Simulations. *J. of*
768 *Hydrometeorology*, **15(3),** 1091-116.
769
770 Rasmussen, R., and Coauthors, 2010: High-Resolution Coupled Climate Runoff Simulations of
771 Seasonal Snowfall over Colorado: A Process Study of Current and Warmer Climate. *J. Climate,*
772 **24**, 3015 – 3048.
773
774 Rodell, M., E. B. McWilliams, J. S. Famiglietti, H. K. Beaudoing, and J. Nigro, 2011: Estimating
775 evapotranspiration using an observation based terrestrial water budget, *Hydrol. Proc.,* **25**, 4082-
776 4092.
777
778 Rodhe, A., 1981: Spring flood meltwater or groundwater? *Nord. Hydrol.,* **12,** 21–30.
779
780 Scanlon, B. R., K.E. Keese, A.L. Flint, L.E. Flint, C.B. Gaye, W.M. Edmunds, I. Simmers
781 (2006). Global synthesis of groundwater recharge in semiarid and arid regions. *Hydrol.*
782 *Proc.,* **20**, 3335–3370 .
783
784 Scanlon, B.R., K. Keese, N. Bonal, N. Deeds, V. Kelley, and M. Litvak, 2005:
785 Evapotranspiration estimates with emphasis on groundwater evapotranspiration in Texas: Austin,
786 Texas Water Development Board, 123 p.
787
788 Schaake, J. C., V. I. Koren, Q.-Y. Duan, K. E. Mitchell, and F. Chen (1996). Simple water
789 balance model for estimating runoff at different spatial and temporal scales. *J. Geophys. Res.,*
790 **101**, 7461–7475, doi:10.1029/95JD02892.
791
792 Siebert, S. J. Burke, J. M. Faures, K. Frenken, J. Hoogeveen, P. Do¨ll, and F. T. Portmann,
793 2010: Groundwater use for irrigation -a global inventory. *Hydrol. Earth Syst. Sci.,***14**, 1863-
794 1880.
795
796 Sleeter, B.M., T. S. Wilson, and W. Acevedo, eds., 2012: Status and trends of land change in the
797 Western United States—1973 to 2000: U.S. Geological Survey Professional Paper 1794–A, 324
798 p. (Available at http://pubs.usgs.gov/pp/1794/a/.)
799
800 Soller, D.R., 1992: Map showing the thickness and character of Quaternary sediments in the
801 glaciated United States east of the Rocky Mountains: U.S. Geological Survey Bulletin 1921, 54
802 p.
803
804 Stonestrom, D. A., J. Constantz, T. P. A Ferré, and S. A. Leake, eds., 2007: Groundwater
805 recharge in the arid and semiarid southwestern United States: U.S. Geological Survey
806 Professional Paper 1703, Reston, VA, 414 p
807
808 Syed, T.H., J.S. Famiglietti, M. Rodell, J.L. Chen, and C.R. Wilson (2008). Analysis of
809 terrestrial water storage changes from GRACE and GLDAS, *Wat. Resour. Res.,* **44**, W02433,
810

Taylor, R. G. and Coauthors, 2013: Ground water and climate change. *Nature Climate Change,* **3(4),** 322-329.

Wahl, K. L., and T.L. Wahl, 1988: Effects of regional ground-water declines on streamflows in the Oklahoma Panhandle, in Proceedings of Symposium on Water-Use Data for Water Resources Management: Tucson, Arizona, American Water Resources Association, p. 239-249, information available on the World Wide Web, accessed August 12, 2015, at URL http://www.usbr.gov/tsc/hydlab/twahl/bfi/bfi_beaver_river.pdf

Wahl, K. L., and T. L. Wahl, 1995: Determining the flow of Comal Springs at New Braunfels, Texas, in Proceedings of Texas Water '95, August 16-17, 1995, San Antonio, Texas: American Society of Civil Engineers, p. 77-86, information available on the World Wide Web, accessed August 12, 2015, at URL http://www.usbr.gov/tsc/hydlab/pubs/PAP/PAP-0708.pdf

Wang, F., L. Wang, T. Koike, H. Zhou, K. Yang, A. Wang, and W. Li, 2011: Evaluation and application of a fine-resolution global data set in a semiarid mesoscale river basin with a distributed biosphere hydrological model. *J. Geophys. Res.,* **116**, D21108, doi:10.1029/2011JD015990.

Wei H., Y. Xia, K. E. Mitchell, and M. B. Ek, 2013: Improvement of the Noah land surface model for warm season processes: evaluation of water and energy flux simulation. *Hydrological Processes*., **27,** 297-303, DOI: 10.1002/hyp.9214.

Wolock, D.M., 2003a: Estimated mean annual natural ground-water recharge in the conterminous United States. U.S. Geological Survey Open-File Report 03-311 Reston, VA. http://water.usgs.gov/lookup/getspatial?rech48grd

Wolock, D.M., 2003b: Base-flow index grid for the conterminous United States: U.S. Geological Survey Open-File Report 03-263, digital dataset, available on the World Wide Web, accessed July 8, 2003, at URL http://water.usgs.gov/lookup/getspatial?bfi48grd

Wolock, D.M., 2003c: Flow characteristics at U.S. Geological Survey streamgages in the conterminous United States: U.S. Geological Survey Open-File Report 03-146, digital dataset, available on the World Wide Web, accessed June 30, 2003, at URL http://water.usgs.gov/lookup/getspatial?qsitesdd

Xia, Y. and Coauthors, 2012a: Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *J. Geophys. Res.,* **117**, D03109, doi:10.1029/2011JD016048.

Xia, Y. and Coauthors, 2012b: Continental-scale water and energy flux analysis and validation for North American Land Data Assimilation System project phase 2 (NLDAS-2): 2. Validation of model-simulated streamflow, *J. Geophys. Res.,* **117**, D03110, doi:10.1029/2011JD016051.

Xia, Y. ad Coauthors, 2013: Validation of Noah-simulated soil temperature in the North American Land Data Assimilation System phase 2. *J. Appl. Meteor. Climatol.,* **52**, 455–471

Xia, Y., J. Sheffield, M. B. Ek, J. Dong, N. Chaney, H. Wei, J. Meng, and E. F. Wood, 2014: Evaluation of multi-model simulated soil moisture in NLDAS-2. *J. Hydrol.,* **512**, 107–125, doi:10.1016/j.jhydrol.2014.02.027.

Xia, Y., M. Ek, Y. Wu, T. Ford, and S. Quiring, 2015a: Comparison of NLDAS-2 Simulated and NASMD Observed Daily Soil Moisture. Part I: Comparison and Analysis. *J. Hydrometeor.* doi:10.1175/JHM-D-14-0096.1, in press.

Xia, Y., M. Ek, Y. Wu, T. Ford, and S. Quiring, 2015b: Comparison of NLDAS-2 Simulated and NASMD Observed Daily Soil Moisture. Part II: Impact of Soil Texture Classification and Vegetation Type Mismatches. *J. Hydrometeor*. doi:10.1175/JHM-D-14-0097.1, in press.

Xue, Y., P. J. Sellers, J. L. Kinter and J. Shukla, 1991: A simplified biosphere model for global climate studies. *J. Climate,* **4**, 345–364.

Table 1: Physical characteristics of study basins

Aquifers	Area (sq. km)	P (mm/yr)	Recharge (mm/yr)	Aquifer material
High Plains(HP)	451,000	535	48	unconsolidated, poorly sorted clay, silt, sand and gravel underlain by bedrock
NHP	250,000	548	73.7	same as HP
CHP	125,000	545	33.7	same as HP
SHP	76,000	472	27.9	same as HP
Central Valley	52,000	650	315.4	sand and gravel
Death Valley	45,300	185	2.8	carbonate and volcanic rock, and alluvium
Colombia	114,000	442	116.7	basalt, sand and gravel
San Pedro	7,560	371	6.5	sand and gravel
Spokane	2,100	689	300.0	sand and gravel
Williston	102,400	382	4.7	sand and gravel

Table 2: Basic differences in LSMs used in this study

	Mosaic (1D)	Noah (1D)	VIC (1D)
Run time step	15 min	15 min	1 hour
Soil Layer	3	4	3
Soli layer depths	10, 30, 160 cm (constant)	10, 30, 60, 100 cm (constant)	10 cm, variable, variable (variable)
Tiling :Vegetation	Y	N	Y
Tiling: Elevation	N	N	Y
Snow Layers	1	1	2
Soil temperature profile	N	Y	Y
Drainage	Y (linear)	Y(linear)	Y(non-linear)
Soil water: vertical diffusion	Y	Y	N
Rooting depth	40 cm (constant)	100 cm (constant) expect forest (down to 200 cm)	Variable (down to 200 cm)
Rooting density	constant	constant	exponential
Canopy capacity	0-1.6 mm	0.5 mm	0.1-1 mm
Convective P input	Y	N	N
PET	Input	Calculates itself	Input
Diurnal Albedo	Y	N	N

*Y:Yes; N:No

Table 3: Pattern correlation (Pearson's r) matrix on ET estimates among LSMs, model ensemble mean (ENS) and MODIS-ET

	Mosaic	VIC	Noah	ENS	MODIS
Mosaic	1.00	0.89*	0.86*	0.97*	0.87*
VIC	0.89*	1.00	0.74*	0.86*	0.77*
Noah	0.86*	0.74*	1.00	0.92*	0.75*
ENS	0.97*	0.86*	0.92*	1.00	0.93*
MODIS	0.87*	0.77*	0.75*	0.93*	1.00

*Statistically significant ($p<0.05$)

Table 4: Water balance comparison between models for study basins and western US

		Mosaic			Noah			VIC		
	P (mm)	ET (mm)	SR (mm)	Recharge (mm)	ET (mm)	SR (mm)	Recharge (mm)	ET (mm)	SR (mm)	Recharge (mm)
Death Valley	185	176.3	7.0	1.6	119.9	12.1	52.8	151.8	10.4	22.5
Colombia	442	403.5	25.9	14.3	296.4	22.9	124.7	243.7	58.1	140.6
Williston	382	375.4	5.2	0.5	332.7	33.1	12.4	310.8	36.3	35.2
San Pedro	371	367.7	2.8	0.2	339.8	14.0	16.8	318.3	27.5	24.9
SHP	472	456.2	7.8	7.1	409.3	28.6	33.0	395.8	32.8	43.3
NHP	548	524.5	9.7	12.8	433.4	39.6	73.9	463.9	41.5	42.4
CHP	545	525.2	9.1	11.0	454.3	34.8	55.6	427.7	26.8	90.9
Central Valley	650	441.6	111.9	97.6	313.1	72.4	266.7	351.5	97.3	201.7
Spokane Valley	689	597.1	71.8	20.7	345.8	53.8	289.6	402.1	118.1	168.6
SLV	488	452.0	24.2	10.3	373.9	33.6	79.1	365.7	41.2	81.4
Western US	561	472.4	43.2	45.5	360.2	61.6	139.1	367.3	66.2	127.8

Table 5: Water balance comparison (in %) between models for study basins and western US

		Mosaic			Noah			VIC			Literature
	P (mm)	ET (%)	SR (%)	Recharge (%)	ET (%)	SR (%)	Recharge (%)	ET (%)	SR (%)	Recharge (%)	Recharge (%)
Death Valley	185	95.4	3.8	0.9	64.9	6.5	28.6	82.2	5.6	12.2	1.5
Colombia	442	91.2	5.9	3.2	67.0	5.2	28.2	55.1	13.1	31.8	26.4
Williston	382	98.2	1.4	0.1	87.0	8.7	3.2	81.3	9.5	9.2	1.2
San Pedro	371	99.2	0.7	0.01	91.7	3.8	4.5	85.9	7.4	6.7	1.7
SHP	472	96.7	1.7	1.5	86.7	6.1	7.0	83.9	6.9	9.2	5.9
NHP	548	95.7	1.8	2.3	79.1	7.2	13.5	84.7	7.6	7.7	13.4
CHP	545	96.3	1.7	2.0	83.3	6.4	10.2	78.4	4.9	16.7	6.2
Central Valley	650	67.9	17.2	15.0	48.1	11.1	41.0	54.0	15.0	31.0	48.5
Spokane Valley	689	86.7	10.4	3.0	50.2	7.8	42.0	58.4	17.1	24.5	43.5
SLV	488	92.6	5.0	2.1	76.6	6.9	16.2	74.9	8.4	16.7	41.6
Western US	561	84.1	7.7	8.1	64.2	11.0	24.8	65.4	11.8	22.8	N/A

Table 6: Pattern correlation (Pearson's r) matrix on recharge estimates among LSMs, model ensemble mean (ENS) and BFI-recharge

	Mosaic	VIC	Noah	ENS	BFI
Mosaic	1.00	0.85*	0.91*	0.95*	0.74*
VIC	0.85*	1.00	0.91*	0.95*	0.77*
Noah	0.91*	0.91*	1.00	0.98*	0.86*
ENS	0.95*	0.95*	0.98*	1.00	0.83*
BFI	0.74*	0.77*	0.86*	0.83*	1.00

*Statistically significant ($p<0.05$)

Table 7: K-S test for comparing distribution of recharge

K-S Test: H0: Samples are drawn from the same distribution

	Dmax				
	Mosaic	VIC	Noah	ENS	BFI
Mosaic	0	0.66	0.65	0.68	0.61
VIC	0.66	0	0.046	0.12	0.37
Noah	0.65	0.046	0	0.11	0.38
ENS	0.68	0.12	0.11	0	0.34
BFI	0.61	0.37	0.38	0.34	0
	p-value				
Mosaic	1	< 0.0001	< 0.0001	< 0.0001	< 0.0001
VIC	< 0.0001	1	< 0.0001	< 0.0001	< 0.0001
Noah	< 0.0001	< 0.0001	1	< 0.0001	< 0.0001
ENS	< 0.0001	< 0.0001	< 0.0001	1	< 0.0001
BFI	< 0.0001	< 0.0001	< 0.0001	< 0.0001	1

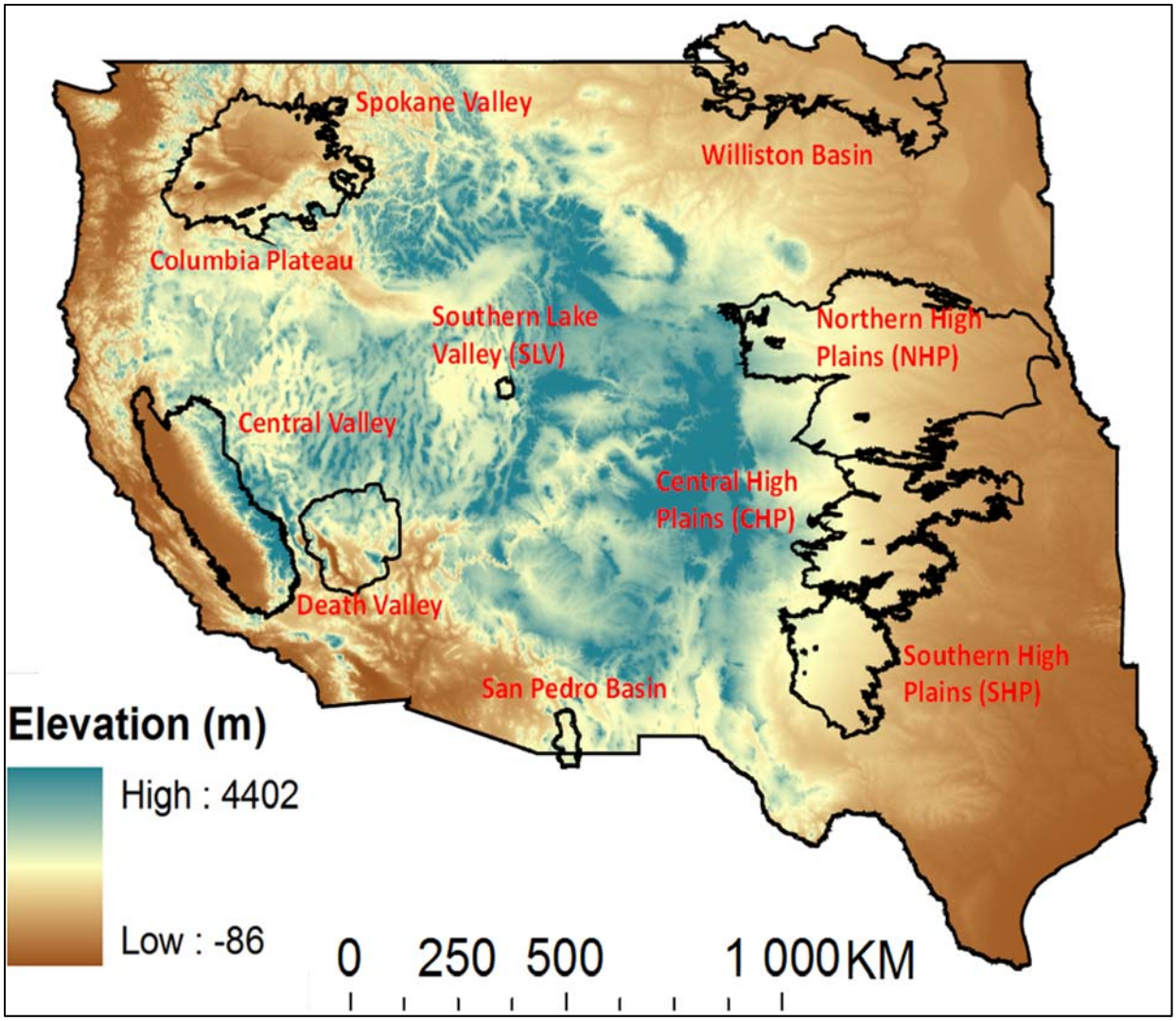

Fig. 1: Study Region shown with elevation and study basins

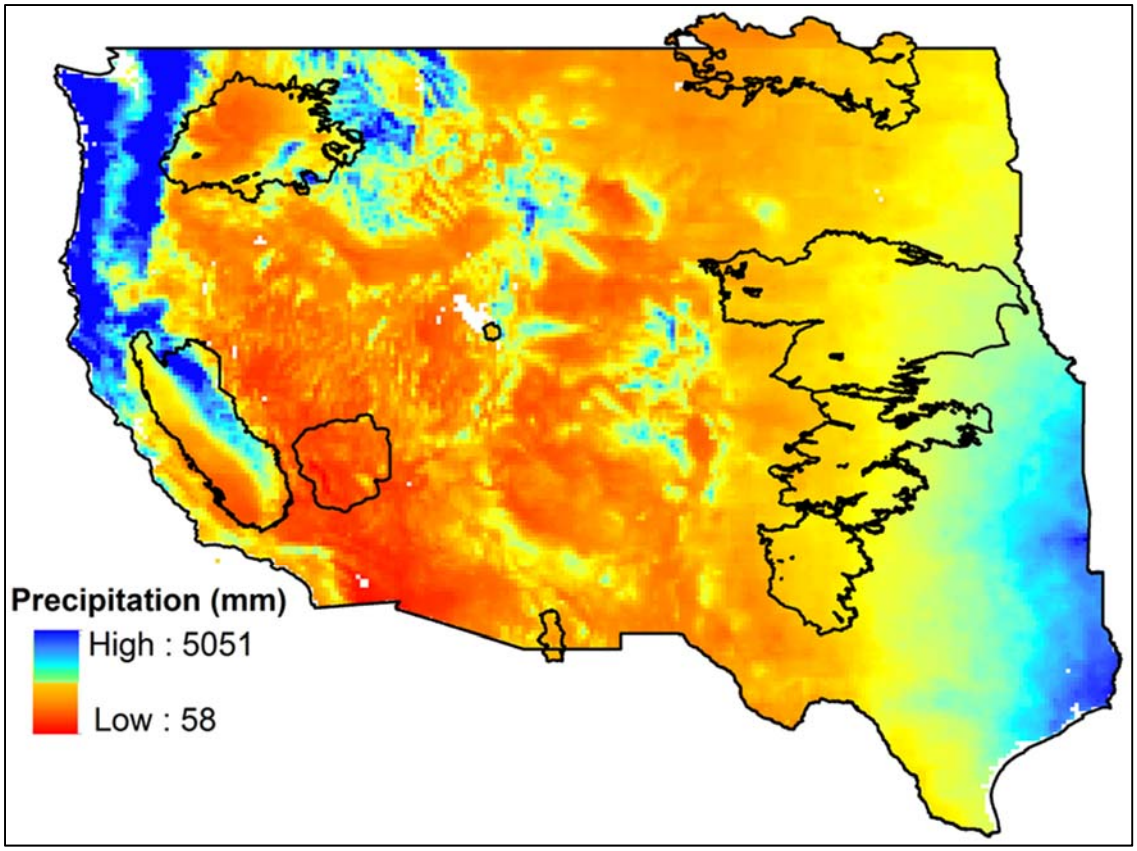

Fig. 2: Average annual precipitation (1981-2010) across the western US

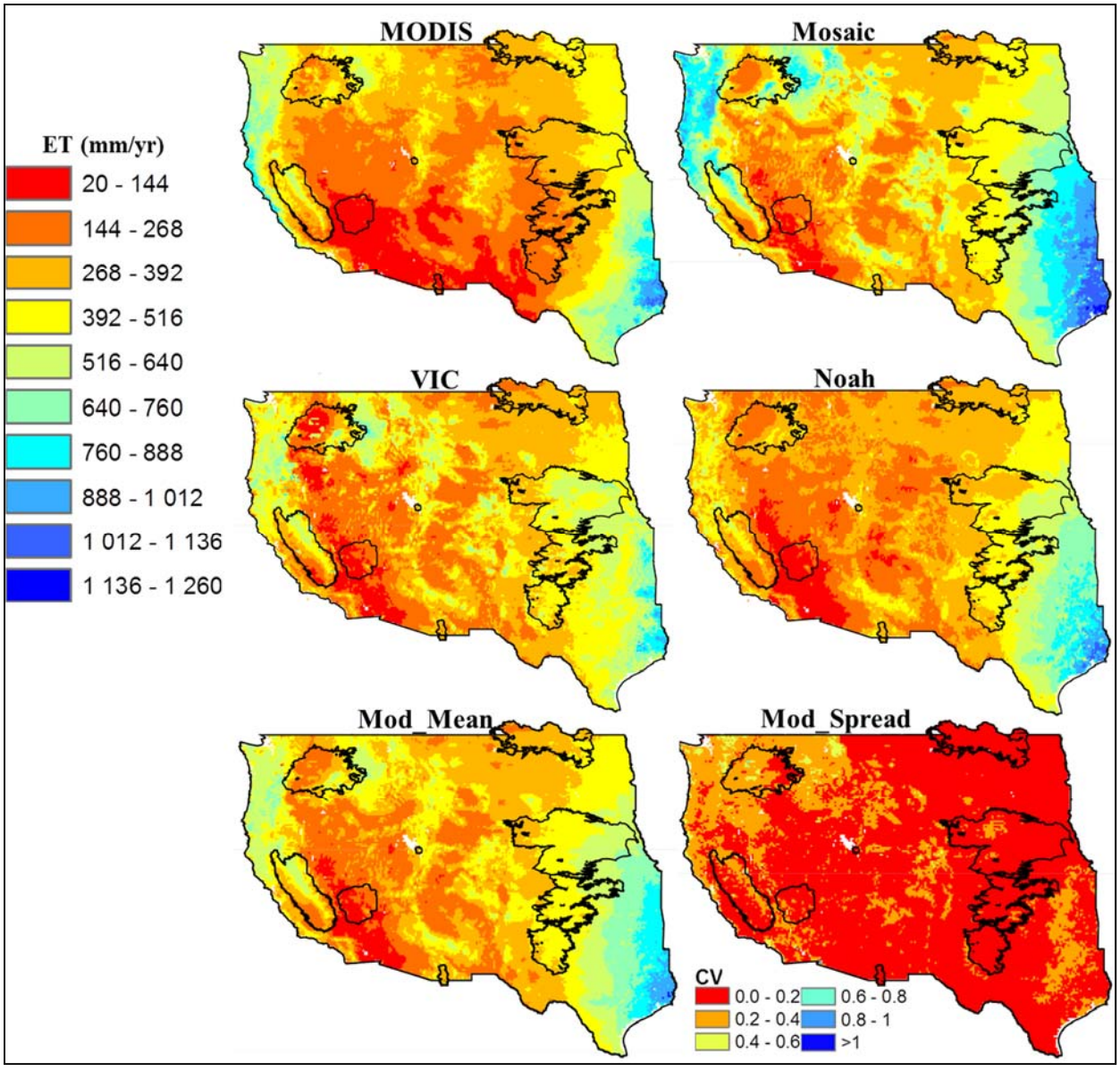

Fig. 3: Annual average ET estimates (2000-2010) from LSMs, MODIS, model ensemble mean and model spread

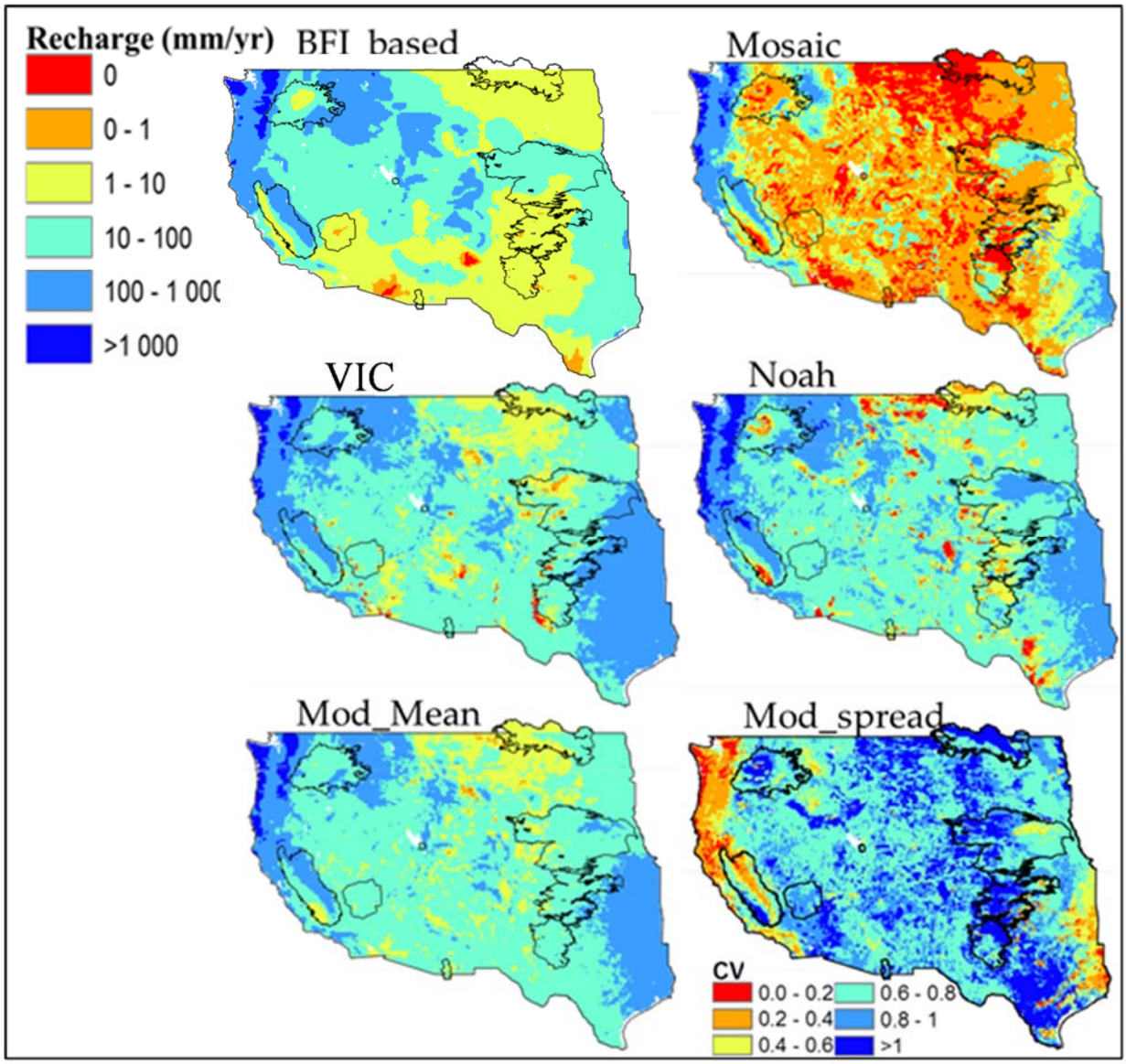

Fig. 4: Average annual recharge estimates (1981-2010) from LSMs, BFI-based recharge, model ensemble mean and model spread

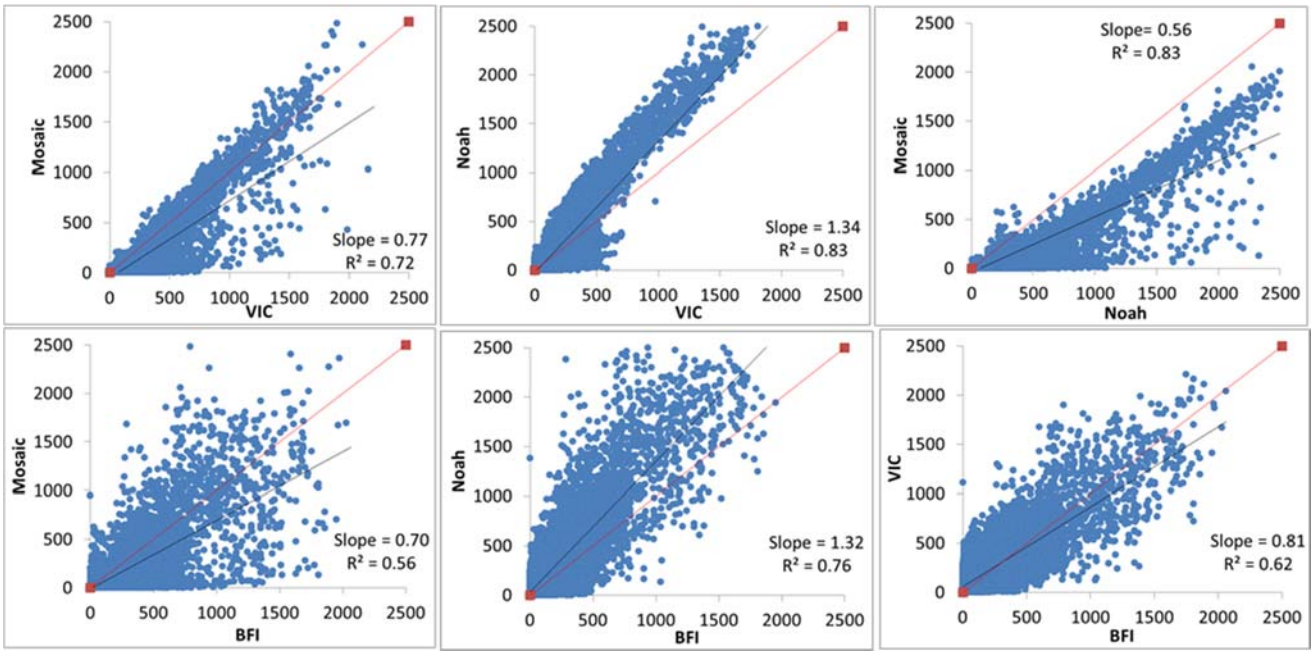

Fig. 5: Scatter plots of recharge between models and models vs BFI-based recharge

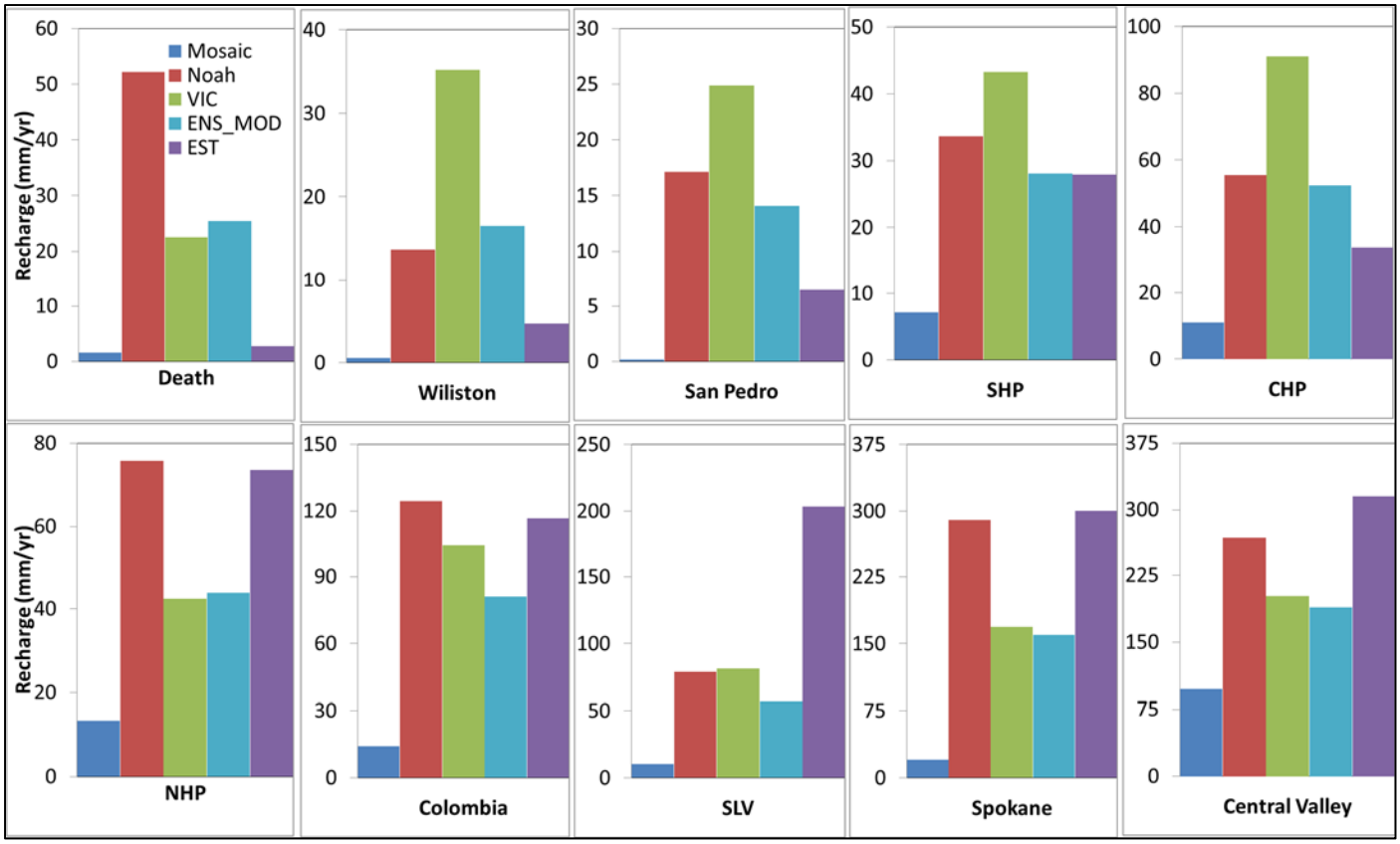

Fig. 6: Comparison of model and ensemble model mean (ENS_MOD) recharge estimates with literature estimates (EST) in 10 study basins

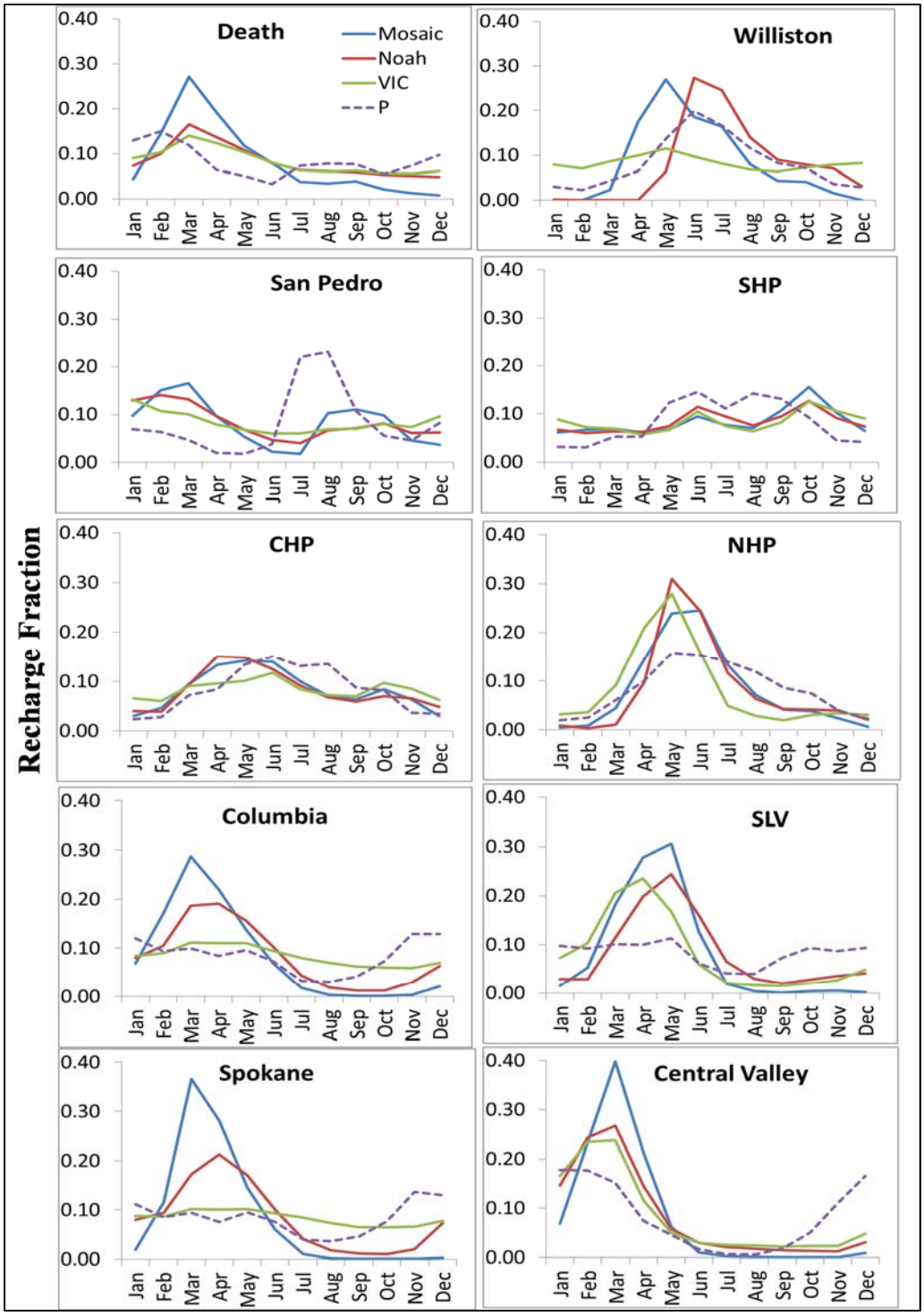

Fig. 7: Seasonlity of recharge based on LSMs and P based on NLDAS-2 data